PB-243 633

ESTABLISHMENT OF A NON-EPA USER SYSTEM FOR STATE IMPLEMENTATION PLANS

PEDCo-Environmental Specialists, Incorporated

Prepared for:

Environmental Protection Agency

January 1975

DISTRIBUTED BY:

NTIS

National Technical Information Service
U. S. DEPARTMENT OF COMMERCE

218134

EPA-450/3-75-049

PB 243 633

JANUARY 1975

ESTABLISHMENT OF A NON-EPA USER SYSTEM FOR STATE IMPLEMENTATION PLANS

Reproduced by
NATIONAL TECHNICAL
INFORMATION SERVICE
Springfield, Va. 22151

U.S. ENVIRONMENTAL PROTECTION AGENCY
Office of Air and Waste Management
Office of Air Quality Planning and Standards
Research Triangle Park, North Carolina 27711

15

PB243633

TECHNICAL REPORT DATA
(Please read Instructions on the reverse before completing)

1. REPORT NO.	2.	3. RECIPIENT'S ACCESSION NO.
EPA-450/3-75-049		

4. TITLE AND SUBTITLE	5. REPORT DATE
Establishment of a Non-EPA User System for State Implementation Plans	January 1975
	6. PERFORMING ORGANIZATION CODE
7. AUTHOR(S)	8. PERFORMING ORGANIZATION REPORT NO.
PEDCo Environmental Specialists, Inc	
9. PERFORMING ORGANIZATION NAME AND ADDRESS	10. PROGRAM ELEMENT NO.
PEDCo Environmental Specialists, Inc Suite 13, - Atkinson Square Cincinnati, Ohio 45246	11. CONTRACT/GRANT NO. 68-02-1001
12. SPONSORING AGENCY NAME AND ADDRESS	13. TYPE OF REPORT AND PERIOD COVERED
U.S. Environmental Protection Agency Office of Air Quality Planning and Standards Research Triangle Park, North Carolina 27711	Final Report 14. SPONSORING AGENCY CODE

15. SUPPLEMENTARY NOTES

16. ABSTRACT

This report presents the results of a survey conducted among selected state air pollution control agencies to determine their current practices and projected needs related to accessing U. S. Environmental Protection Agency data bases. Alternative methods for allowing non-EPA users to use the data bases were introduced. A preliminary cost survey was conducted for a projected method for allowing state agencies to have direct access to the data bases.

This is a preliminary analysis of expected costs for operating the Comprehensive Data Handling System on a centralized computer accessed through remote terminals located in state air pollution control agency offices. The analysis was compiled from data obtained in informal discussions with four (4) computer centers that might be expected to be candidate contractors for such a system. System usage inputs were obtained from Task I of this contract.

17. KEY WORDS AND DOCUMENT ANALYSIS

a. DESCRIPTORS	b. IDENTIFIERS/OPEN ENDED TERMS	c. COSATI Field/Group
EPA Data Bases NEDS SAROAD Emissions Data Air Quality Data Compliance and Enforcement Data SIP Regulation Data		

18. DISTRIBUTION STATEMENT	19. SECURITY CLASS (This Report)	21. NO. OF PAGES
Release Unlimited	Unclassified	75
	20. SECURITY CLASS (This page)	22. PRICE
	Unclassified	4.25

EPA Form 2220-1 (9-73)

This report is issued by the Environmental Protection Agency to report technical data of interest to a limited number of readers. Copies are available free of charge to Federal employees, current contractors and grantees, and nonprofit organizations - as supplies permit - from the Air Pollution Technical Information Center, Environmental Protection Agency, Research Triangle Park, North Carolina 27711; or, for a fee, from the National Technical Information Service, 5285 Port Royal Road, Springfield, Virginia 22161.

This report was furnished to the Environmental Protection Agency by PEDCo-Environmental Specialists, Inc., Cincinnati, Ohio 45246, in fulfillment of Contract No. 68-02-1001. The contents of this report are reproduced herein as received from PEDCo-Environmental Specialists, Inc. The opinions, findings, and conclusions expressed are those of the author and not necessarily those of the Environmental Protection Agency. Mention of company or product names is not to be considered as an endorsement by the Environmental Protection Agency.

Publication No. EPA-450/3-75-049

ACKNOWLEDGMENTS

Task I of this report was prepared by PEDCo-Environmental Specialists under the direction of Mr. Charles E. Zimmer. Mr. David W. Armentrout served as Project Manager. We would like to express our appreciation to the following organizations for participating in this survey and for providing technical information:

- Arizona Division of Air Pollution Control
- Argonne National Laboratory
- Cleveland, Ohio, Division of Air Pollution Control
- Connecticut Air Compliance Section
- Indiana Division of Air Pollution Control
- Lawrence Livermore Laboratory
- Maryland Bureau of Air Quality Control
- New Mexico Air Quality Section
- Ohio Environmental Protection Agency
- Washington State Department of Ecology
- District of Columbia Bureau of Air and Water Pollution Control
- United States Department of Housing and Urban Development

Mr. Gerald Nehls, National Air Data Branch, served as Project Officer for the Environmental Protection Agency. We wish to thank Mr. Nehls for his valuable assistance.

Task II of this report was prepared for PEDCo-Environmental Specialists by PRC Systems Sciences Company under Subcontract No. 63-74. We wish to express our appreciation for the contributions of Mr. Charles Bloomquist, PRC Systems Sciences Company, who conducted the surveys to gather cost data, made the cost projections, and wrote the report. The report was subsequently edited by PEDCo-Environmental Specialists.

ABSTRACT

This report presents the results of a survey conducted among selected state air pollution control agencies to determine their current practices and projected needs related to accessing U.S. Environmental Protection Agency data bases. Alternative methods for allowing non-EPA users to use the data bases were introduced. A preliminary cost survey was conducted for a projected method for allowing state agencies to have direct access to the data bases.

PREFACE

This survey was conducted to identify requirements and quantify costs for direct state access to air data bases. Nine state agencies were surveyed and results were extrapolated to cover the nation by the "State Air Data Information Survey" EPA-450/3-74-001.

The states to be surveyed were determined by examining their data handling capability to have a mix of manual and computerized systems. The regional office AEROS contacts also participated in the selection.

The requirements for direct access to air data bases were not overwhelming. Some states would use any system which is cost-effective. Others would not use a system even if it were cost effective. The costs of providing states with direct access are estimated in this report. No attempt was made to estimate state usage to fulfill their own requests or requirements other than for federal reporting.

As a result of this survey the National Air Data Branch will concentrate on reducing turnaround times for updating data and retrieving data and implementing the Comprehensive Data Handling System for interested state agencies. We will not plan to provide direct state terminal access to the AEROS System.

An EPA user survey, soon to be released will identify user requirements and requested enhancements to the AEROS systems. These will be developed as resources permit.

Gerald Mehls

TABLE OF CONTENTS

TASK I. SURVEY OF STATE AND LOCAL AGENCIES	Page
1.0 EXECUTIVE SUMMARY	1
1.1 Background	1
1.2 Approach	1
1.3 Survey Results	2
1.3.1 State Agency Applications for EPA Data Bases	2
1.3.2 Potential Users of EPA Data Bases	2
1.3.3 Alternative Systems for Accessing NEDS/SAROAD Data	3
2.0 TECHNICAL REPORT	5
2.1 Introduction	5
2.1.1 Background	5
2.1.2 Purpose and Scope of the Survey	5
2.2 Survey Procedures	6
2.2.1 Background Preparation	6
2.2.2 Content of the Survey	7
2.2.3 Organizations Surveyed	9
2.3 Survey Results	9
2.3.1 Interpretive Restrictions	9
2.3.2 Emissions Data	11
2.3.3 Air Quality Data	14
2.3.4 Compliance and Enforcement Data	16
2.3.5 SIP Regulation Data	16
2.3.6 Selected Non-EPA Data Tapes	17
3.0 CONCLUSIONS	18
3.1 User Classification	18
3.1.1 Basic Data Handling Systems	18
3.1.2 Intermediate Data Handling Systems	18
3.1.3 Advanced Data Handling Systems	19
3.2 Candidate Users for Alternative Systems	19
3.3 Projected Activity for Alternative Systems	20
3.4 Alternative Systems for NEDS/SAROAD Access	25

TASK II. COST ANALYSIS

1.0 EXECUTIVE SUMMARY 28

1.1 System Overview 28
1.2 conclusions 30

2.0 TECHNICAL REPORT 33

2.1 System Overview 34
2.2 Candidate Network 38

2.2.1 Assumptions 38
2.2.2 Network Evaluation Input 41
2.2.3 Cost Review 61
2.2.4 Evaluation of Cost Survey Results 63

Technical Report Data and Abstract 66

LIST OF FIGURES

Figure Page

1. Research Triangle Computer Center Input 58

LIST OF TABLES

Table		Page
1.	Anticipated Use of NEDS/SAROAD	3
2.	NEDS Reports for Meeting Minimum Emissions Reporting Requirements of State Agencies	12
3.	SAROAD Reports for Meeting Minimum Air Quality Reporting Requirements of State Agencies	15
4.	Estimates of Computer Status for State E.I. (Emission Inventory) and A.Q. (Air Quality) Data Systems	20
5.	Minimum Summary Frequencies for E.I. and A.Q. Data in State Agencies	21
6.	Annual Range of Inquiries for E.I. and A.Q. Data Reports	21
7.	Minimum E.I. Report Volume for a Range of 7-29 States	22
8.	Minimum A.Q. Report Volume for a Range of 3-18 States	22
9.	Update Frequencies for Computerized E.I. Data Handling Systems	23
10.	Update Frequencies for Computerized A.Q. Data Handling Systems	23
11.	Volume of Update Data for E.I. Systems	24
12.	Reporting Stations Volume of Update Data for A.Q. Systems	25
13.	Projected Stations Volume of Update Data for A.Q. Systems	25
14.	Network Benefit Comparison	31
15.	Network First-Year Cost Comparison	32
16.	States with Manual Data Systems	35
17.	Annual Data Volumes	40
18.	Typical Report Processing Times	42
19.	Terminal Network Cost Estimates	43
20.	BCS Terminal Network Cost Estimates	45
21.	EPA/OSI Terminal Network Cost Estimates	50
22.	GSA INFONET Terminal Network Cost Estimates	53
23.	Length of Typical GSA INFONET Telephone Toll Call	54
24.	RTCC Terminal Network Cost Estimates	59
25.	Typical Terminal Lease Charges	60
26.	Typical CPU Time for CDHS Installation on an IBM 360/50	60
27.	Network First-Year Cost Comparison	62

TASK I. SURVEY OF STATE AND LOCAL AGENCIES

1.0 EXECUTIVE SUMMARY

1.1 BACKGROUND

The Environmental Protection Agency (EPA) has developed extensive computer-oriented systems for the storage and retrieval of data on air quality measurements and air contaminant emissions. The EPA system for handling air quality data is the Storage and Retrieval of Aerometric Data (SAROAD) and the system for handling emissions data is the National Emissions Data System (NEDS). EPA is developing systems to create national data bases for state and local emission regulations, source tests, and hazardous pollutant emissions. Each state and local control agency has access to software included in EPA's Comprehensive Data Handling System (CDHS), which gives state agencies the ability to establish their own computerized air pollution data bases.

The purposes of the work performed under this contract were to:

- Survey state and local air pollution control agencies and selected Federal organizations to determine the extent to which their data requirements are fulfilled by the various EPA data bases.
- In the event that existing EPA data systems do not fulfill state and local agency requirements, perform a cost/effectiveness analysis of alternative data systems required to meet these needs.

1.2 APPROACH

A two-man interview team visited selected state and

Federal agencies and reviewed the data systems now being used. Personnel of each agency were questioned regarding the reports and data analyses they prepare and the ways in which management uses this information for decision making and planning. To assist the interviewers in maintaining consistency, a survey questionnaire was developed. Agencies to be included in the study were selected in consort with NEDS/SAROAD coordinators in EPA Regional Offices.

1.3 SURVEY RESULTS

The results of this survey are summarized here in terms of current and projected use of the NEDS and SAROAD data bases by state agencies. These results provide the basis for the cost/effectiveness study to be performed in Task II. Information obtained from local agencies and from the Federal research and planning organizations is not included in these results and projections but is analyzed separately in Attachment 1 (provided to the Project Officer).

1.3.1 State Agency Applications for EPA Data Bases

All of the agencies interviewed were familiar with NEDS and SAROAD, but none were aware of the other data bases. None of the agencies were familiar with all the report and retrieval options of NEDS and SAROAD. All of the states currently using manual or marginally computerized systems for emissions and air quality data could benefit from increased use of NEDS and SAROAD.

The agencies indicated interest in the other data bases discussed, but did not identify specific applications or projected access frequencies. All agencies indicated that access to these other EPA data bases by batch request through EPA Regional Offices would be adequate.

1.3.2 Potential Users of EPA Data Bases

State agencies interviewed during this survey can be classified as basic, intermediate, or advanced in terms of data-handling systems. The basic category includes states

with totally manual systems. The intermediate category includes states having some computerization of emissions data or air quality data, but usually not both. The advanced category includes states with both types of data computerized.

The potential users of any alternative system for accessing NEDS/SAROAD data are primarily those states classified as basic or intermediate. The numbers of potential users are expressed as ranges, rather than absolutes. Based on the results of EPA's "State Air Data Information Survey," (EPA-450/3-74-001, January 1974) ranges have been calculated for potential users of both NEDS and SAROAD. NEDS users would be anticipated to include from 7 to 29 states. SAROAD users would be anticipated to include from 3 to 18 states. Users of NEDS cannot necessarily be anticipated to be users of SAROAD. Anticipated ranges of output for NEDS and SAROAD are summarized in Table 1.

Table 1. ANTICIPATED USE OF NEDS/SAROAD

	NEDS	SAROAD
Range of user states	7-29	3-18
Average output volume, pages/year/state	2,300	1,900
Total volume range, pages	18,000-65,000	56,000-34,000
Total annual update volume	202-842 (Point sources)	484,000-1,356,000 (Data values)

1.3.3 Alternative Systems for Accessing NEDS/SAROAD Data

Several alternative systems could provide the states classified as basic and intermediate with access to NEDS and SAROAD data. Any system used would require a capability for the users to submit directly to the computer rather than through the EPA Regional Offices, as is the current practice. Four proposed systems provide the basis for the cost analyses to be performed in Task II.

1. Batch Processing Through EPA Regional Offices - This is the mechanism currently used by states requiring NEDS/SAROAD reports. Any state not using this option would meet their data handling requirements by installing CDHS subsystems or by developing systems in-house.
2. Access to NEDS/SAROAD Via Remote Terminal - With this system data from both NEDS and SAROAD would be available from CDHS implemented on a central computer.
3. Access to NEDS Via Remote Terminal - With this system NEDS data would be accessible from the Emissions Inventory Subsystem (EIS) of CDHS implemented on a central computer. States with requirements for air quality data would meet their requirements by requesting SAROAD reports through the EPA Regional Offices, by developing their own systems, or by implementing the Air Quality Data Handling Subsystem (AQDHS) of CDHS on their own computers.
4. Access to SAROAD Via Remote Terminal - With this system, SAROAD data would be accessible from AQDHS implemented on a central computer. States with requirements for emissions data would request NEDS reports through the EPA Regional Offices, or they could develop their own systems or implement EIS on their own computers.

2.0 TECHNICAL REPORT

2.1 INTRODUCTION

2.1.1 Background

The Storage and Retrieval of Aerometric Data System (SAROAD) and the National Emission Data System (NEDS) were developed by EPA to provide information on measurements of air contaminants in the atmosphere and emissions of air contaminants from area and point sources. The data bases provided by these systems are used by state air pollution control agencies and by EPA for testing of control strategies, for evaluation of State Implementation Plans (SIP's), and for SIP report and follow-up activities required by Federal law. They are also being used to provide air quality and emissions data needed for designations of Air Quality Maintenance Areas.

EPA is currently developing a computer-oriented storage and retrieval system to handle the information on air quality regulations contained in the SIP's. Other associated data bases are also available through or are being developed by EPA. Output from SAROAD, NEDS, the SIP's, and the other EPA data bases are available for use by state and local governments, research groups (universities, institutes, etc.), regional planning groups, and industrial-commercial organizations. The EPA data bases are currently accessible through requests made to EPA Regional Offices or made directly to EPA's National Air Data Branch in North Carolina.

2.1.2 Purpose and Scope of the Survey

This survey was conducted to determine to what extent

the EPA systems maintained by the National Air Data Branch currently satisfy the data storage and retrieval requirements of non-EPA users of the systems. To make this determination, personal interviews were conducted to obtain answers to the following questions:

- What are the retrieval requirements, volumes and request frequencies for information contained in each system?
- What are the requirements for using information from multiple data bases to perform an analysis?
- What are the query requirements of the user (as opposed to long summaries and analyses)?
- What requirements (if any) does the user have for direct access to various data bases and software?
- How does the user believe EPA should allow him to access the information and how does he believe this should be implemented?

The major emphasis of the survey was placed on determining requirements of selected state air pollution control agencies. Recommendations for performing cost analyses for development of additional or alternative EPA systems are based primarily on the results of these state surveys.

2.2 SURVEY PROCEDURES

2.2.1 Background Preparation

The first step in preparation for the survey was to review the following air data bases and the data handling systems which are currently available through EPA:

- NEDS - National Emission Data System
- SAROAD - Storage and Retrieval of Aerometric Data
- SOTDAT - Source Test Data System
- EIS - Emission Inventory Subsystem (CDHS)
- AQDHS - Air Quality Data Handling Subsystem (CDHS)
- DQIS - Data Quality Information System
- CDS - Compliance Data System
- EMS - Enforcement Management Subsystem (CDHS)
- HAPEMS - Hazardous Air Pollutants Enforcement Management System

- HATREMS - Hazardous and Trace Emissions System
- SIP Rules and Regulations
- Non-EPA Selected Tapes (Bureau of Census, GSA Property, Form FPC-67, and Federal Facilities Files)

To provide guidance in determining the form and content of the survey format, PEDCo reviewed the "Survey of EPA User Interest for Proposed State Implementation Plan Automated Information System" (EPA-450/3-73-011). The format of the formal survey questionnaire was designed to be compatible with the format of the EPA user survey.

2.2.2 Content of the Survey

The PEDCo survey, known as the Non-EPA Users Survey, was designed to meet several criteria:

- Provide a multipurpose format suitable for obtaining information on the air data requirements of organizations with diverse applications and needs.
- Obtain answers to the five questions outlined in Section 2.1.
- Provide compatibility with the completed "Survey of EPA User Interest for Proposed State Implementation Plan Automated Information System."
- Enable interviewers to spend minimal time discussing areas of little concern to the user or potential user, concentrating, rather, on areas of maximum interest.
- Relate current and projected data requirements and data handling practices to the data bases and data systems currently available through EPA or being developed.

Attachment 2 (provided to the Project Officer) presents a sample questionnaire used in this survey. The questionnaire is divided into five major sections by data type:

- Emissions Data
  - NEDS
  - SOTDAT
  - EIS
  - HATREMS (not shown on the questionnaire)

° Air Quality Data
  SAROAD
  AQDHS
  DQIS
° Compliance and Enforcement Data
  CDS
  EMS
  HAPEMS
° SIP Regulation Data
° Non-EPA Data
  Bureau of Census Tapes
  GSA Property Files
  Form FPC 67
  Federal Facility Files
  INFONET models (not shown on questionnaire)

Each of these major sections of the questionnaire begins with a detailed query regarding the agency's current requirements for various formats, turnaround times, mechanisms for query, and the like. This general, user-oriented inquiry is then followed by questions specific to the EPA data bases and systems related to the major data category under discussion. Inquiry concerning the specific EPA systems was pursued only if the agency indicated need for or interest in those applications. Where appropriate, sample output forms for NEDS and SAROAD were presented and discussed.

Only data bases of concern to a specific organization were discussed in detail. Sample output formats for NEDS and SAROAD were presented and discussed to determine how they related to the organization's requirements. Several questions provided information about the users and their data requirements, current formats, current operating mechanisms, and the applicability of existing EPA data bases and systems. Information concerning query requirements, request volume, desired formats, and desired mechanisms for accessing data bases was directed toward evaluating the requirements for accessing the

data banks through remote terminals. Moreover, this line of questioning provided information with which to relate current requirements to data handling systems now in development but not yet released by EPA.

2.2.3 Organizations Surveyed

Nine state agencies and three Federal organizations were included in the survey; these organizations are listed in the Acknowledgement. The nine state agencies were selected to provide representation of states operating three types of air data handling systems:

- Totally manual data handling systems.
- Static computer systems developed to handle basic reporting requirements.
- Dynamic computer systems developed to handle multiple reporting requirements.

NEDS/SAROAD coordinators in the EPA Regional Offices provided recommendations concerning the state agencies to be included in the survey. As an aid in evaluating the representativeness of the agencies in terms of their data systems and in projecting their needs for direct access to the EPA data bases, PEDCo used EPA's "State Air Data Information Survey" (EPA-450/3-74-001). Since the PEDCo survey entailed only a small sample of agencies, this earlier EPA survey provided the best available means for extrapolating our results to a national scale.

In arranging for the interviews with state agencies, we requested that they include persons representing the major user groups or areas of responsibility (i.e. surveillance monitoring, enforcement, permit issuance). In this way the survey could encompass experience from the working level as well as from the management level of an agency.

2.3 SURVEY RESULTS

2.3.1 Interpretive Restrictions

2.3.1.1 Representativeness of the Survey Sample - The survey sample included nine air pollution control agencies and three Federal research and planning organizations involved with environmentally related activities. Information obtained from

the three Federal organizations provided some insight into the data system requirements of organizations whose function is other than control. Because no previous surveys of project requirements or computer capabilities of such organizations have been reported, we have no basis for extrapolating the information obtained concerning their requirements to those of other organizations of that type. Further comments and recommendations concerning these organizations are given in Attachment 1 (given to the Project Officer).

The survey plan anticipated that information obtained from the state agencies would also represent the requirements of local control agencies. This is not the case. The local control agency structures are diverse in terms of both technical expertise and operational autonomy. Consequently, the requirements of the state agencies for data storage and retrieval systems should not be interpreted as representing the requirements of local agencies. It is emphasized, therefore, that the conclusions and recommendations of this survey are based solely on the requirements of state control agencies.

2.3.1.2 Alternative Federal Systems - EPA has developed the Comprehensive Data Handling System (CDHS) for installation and operation at the state and local levels. This system incorporates three subsystems:

- Emission Inventory Subsystem (EIS)
- Air Quality Data Handling Subsystem (AQDHS)
- Enforcement Management Subsystem (EMS)

The subsystems correspond functionally to NEDS, SAROAD, and CDS (Compliance Data System), respectively. Installation of any or all of the CDHS subsystems by a state agency could significantly reduce any requirement for the agency to access the corresponding EPA data system. Recommendations resulting from this survey are based to a great extent on states that have already installed subsystems of CDHS or have firm commitments to install them.

2.3.1.3 Estimates of Frequency and Volume for Data Retrieval - The state agencies surveyed indicated that they were more aware of their requirements for providing input to NEDS and SAROAD than they were of their options for retrieving reports from those data bases. Most states were familiar with only the "NEDS Point Source Listing," the SAROAD "Site Description Inventory," and the SAROAD "Yearly Report by Quarters." Since the other reports had not been requested, the frequency of requests and volume of data were not discussed. The states that had received the three basic reports indicated that turn-around time for receiving reports through the EPA Regional Offices is too long and that the data received are thus not current enough for their applications. Most states were unaware of ways in which NEDS/SAROAD could help them do their jobs more efficiently.

Since report retrieval estimates were not available from the agencies, the projected system usage figures in this report are based primarily on input and retrieval levels that would allow the agencies to meet their formal reporting requirements to EPA. Estimates of activity levels generated by the internal requirements of the agencies were not attempted.

2.3.2 Emissions Data

2.3.2.1 Applications - The major applications identified for emissions data were for use in air quality modeling and in designation of AQMA's. Requirements for data on emissions from specific point sources are usually met by accessing the state permit file. Many states indicated an interest in area source data, apparently the weakest portion of each state's emission inventory. The applications identified for emissions data did not indicate a need for scheduled accessing of an emission data system to retrieve specific reports.

2.3.2.2 Retrieval Systems - Three states have computerized emission inventory systems. Six states have totally manual emission inventory systems, and two of these have initiated plans to install EIS. All states indicated that they prefer

to contact states directly for any required emissions data from those states. The frequency of such requests is very low. Existing systems, even in the states with computer retrieval, do not meet the query requirements for emissions data. The query types of requests for emissions data are usually fulfilled by review of permit files.

Systems for the storage and retrieval of emission data at the state level receive secondary development consideration. Primary consideration for systems development is given to air quality data.

2.3.2.3 NEDS Requirements - All state agencies interviewed were familiar with NEDS. Their familiarity, however, came primarily from EPA's requirements for states to update NEDS data as part of their semi-annual reporting. The state agencies had no significant experience with receiving and using NEDS reports routinely. One state has never requested NEDS output; the others have received at least one "NEDS Point Source Listing." None were familiar with the other available NEDS output formats.

A review of the formal reports produced by the states indicates that, although their format requirements vary slightly, the formats available from NEDS are generally compatible with, or adaptable to, their requirements. They suggested no format changes or additions. The reports from NEDS that could help most states meet their formal and informal reporting requirements are shown in Table 2, which indicates also the minimum frequency with which each report would be requested.

Table 2. NEDS REPORTS FOR MEETING MINIMUM EMISSIONS REPORTING REQUIREMENTS OF STATE AGENCIES

NEDS report	Minimum annual request frequency
Point source listing	2/state/year
Point source listing (for selected sources only)	14/state/year
Emission summary	2/county/year
Area source listing	1/county/year

The estimates in Table 2 were derived from reviewing the current reporting practices of the nine state agencies included in the survey.

The current selection options for retrieval of each NEDS report format are adequate for all states interviewed. Special interest was expressed in the options for accessing the file by Source Classification Code (SCC), since most applications for emissions data relate to special studies entailing specific fuel types, boiler types, processes, or other source category. No state in the survey has significant requirements for emissions data from adjacent states.

For most of the states, turnaround time of 1 to 2 weeks from time of request to receipt of a NEDS report is adequate. One state requested a 1-day turnaround time. Most states are discouraged with the current mechanism for batch requests through the Regional Offices. They indicated that some turnaround times have been as long as 6 weeks, which they believe to be unreasonable. An additional problem with the NEDS mechanism is that the time lag between submittal of data to EPA and the actual update of the computer file, which can be as long as 6 months, is intolerable. Most states indicated that they could envision more uses for NEDS if they could be confident that NEDS reports represent current conditions. Turnaround time represents the most significant objection to accessing NEDS.

Five states surveyed would be potential users of a remote terminal system for accessing NEDS: two of these indicating that their use would be marginal. All potential users expressed two prerequisites for using a terminal system to access NEDS.

- The system should be cost/effective, i.e. it should be less expensive to operate than EIS or an in-house system.
- An interactive capability must be supplied to allow the user to input emissions data or to update the

data base directly. This capability would eliminate the objection of the time lag currently encountered when data are submitted through the Regional Offices.

Request levels for NEDS were not specified, since no state has made extensive use of the system.

2.3.2.4 Other Data Bases - The other emissions data bases discussed in the survey were SOTDAT (Source Test Data) and HATREMS (Hazardous and Trace Emissions System). HATREMS generated no specific interest. SOTDAT was recognized as a means of obtaining emissions estimates for specific sources. None of the agencies interviewed is engaged in extensive factor development work.

2.3.3 Air Quality Data

2.3.3.1 Applications - The major applications identified for ambient air quality data were for control strategy testing, designation of AQMA's, and urban planning (transportation, land use, etc.). The pollutants of major concern are suspended particulates, SO_2, CO, and photochemical oxidants. One state expressed interest in source-oriented air quality data.

2.3.3.2 Retrieval Systems - Five states have manual air quality data handling systems. Two states have sophisticated computer systems to analyze data transmitted directly from field monitors. Two states have adapted the original version of AQDHS to meet their requirements. All states prefer to contact other states directly to obtain any required air quality data. The frequency of those requests is low.

Systems for storage, analysis, and retrieval of ambient air quality data receive development priority in all states in the survey. The agencies were better prepared to discuss their format requirements for air quality data than for emissions data. Several states expressed interest in developing capabilities for trends analysis and also in computer graphics for display of air quality trends data.

2.3.3.3 SAROAD Requirements - All state agencies interviewed were familiar with SAROAD. Experience with using SAROAD and

acceptance of SAROAD reporting requirements is more widespread than for NEDS. SAROAD users have more confidence in the quality of the data than they have in the quality of NEDS emissions data. As with NEDS, however, SAROAD users are generally not familiar with the report formats available from SAROAD. The formats most familiar are the "Quarterly Frequency Distribution," and the "Yearly Report by Quarters."

A review of the formal reports produced by the states indicates that summaries of air quality data are published periodically from monthly to annually. All of the available SAROAD formats are compatible with or readily adaptable to the formal report requirements of the states. SAROAD reports that could help most states meet their formal and informal reporting requirements are shown in Table 3, which also indicates the minimum frequency with which each report would be requested.

Table 3. SAROAD REPORTS FOR MEETING MINIMUM AIR QUALITY REPORTING REQUIREMENTS OF STATE AGENCIES

SAROAD report	Minumum annual reports
Raw data listing	64/sensor
Quarterly reports	4/sensor
Quarterly frequency distribution	4/sensor
Yearly frequency distribution	1/sensor
Yearly report by quarters	1/sensor

The selection options for retrieval of SAROAD report formats are adequate.

Most agencies indicated that turnaround time of 1 week from time of request to receipt of a SAROAD report is adequate. One state requested 1-day turnaround time. The states expressed the same concerns with long turnaround for SAROAD processing through the Regional Offices as they expressed for NEDS processing. Lag time between data submittal and computer update is more critical for SAROAD than for NEDS, because more

data are submitted more frequently. Moreover, air quality data have the potential of being used for analyses related to air alerts.

Seven states indicated that they would be potential users of a remote terminal system for accessing SAROAD: two of these would be marginal users. All potential users expressed two prerequisites for using a terminal system to access SAROAD:

- The system should be cost/effective, i.e. it should be less expensive to operate than AQDHS or an in-house system.
- An interactive capability must be supplied to allow the user to input air quality data directly, or to update the data base directly. This capability would eliminate the objection of the time lag currently encountered when data are summarized through the Regional Offices.

Request levels for SAROAD reports were not specified in the interviews, since no state has made extensive use of the system.

2.3.3.4 Other Data Bases - The other data base discussed for air quality data was DQIS for quality control data. No interest was expressed in this system.

2.3.4 Compliance and Enforcement Data

The three systems discussed for compliance and enforcement data were EMS, CDS, and HAPEMS. No one expressed interest in HAPEMS. Only three states expressed interest in EMS or CDS, and that interest is only in the passive segments of the systems. One state has an operational system for tracking compliance information for hazardous pollutant emission sources. None of the other states could outline their requirements for compliance data.

2.3.5 SIP Regulation Data

Most states commenting on the utility of accessing regulations data indicated that they have been satisfied to request regulations data directly from the state involved. No states cited past experience with reviewing regulations from other states. All states indicated that batch requests through EPA

Regional Offices for information in the developmental SIP Regulations data base would be adequate.

2.3.6 Selected Non-EPA Data Tapes

The selected Non-EPA data tapes discussed were for Federal Facility Files, FPC Form - 67, Bureau of Census Population Data, and GSA Property Files. No interest was expressed for accessing any of these data bases.

The INFONET system for air quality modeling was also discussed. Some states indicated a desire to obtain the model tapes and install them on their own computers. One state agency has had a formal system presentation and price proposal from the contracting agency handling INFONET, but indicated that the system is not cost-justified for them.

3.0 CONCLUSIONS

3.1 USER CLASSIFICATION

The state agencies covered in this survey are classified as basic, intermediate, or advanced in terms of their current capabilities for storage, analysis, and retrieval of air pollution data. Among the states included in the survey, three are categorized as basic, four as intermediate, and two as advanced. The important data handling characteristics of state agencies in each category are summarized in the following paragraphs. An extrapolation for categorizing all state or equivalent agencies is explained in Section 3.2.

3.1.1 Basic Data Handling Systems

Agencies in this category are characterized as having no operational systems for computerized storage and retrieval of air pollution data. All calculations, trends analyses, and report summaries are performed manually. These states have made no commitments to convert to computer-based systems.

NEDS and SAROAD could fulfill the data handling requirements of these agencies. The reporting formats available from NEDS and SAROAD represent a potentially significant improvement in the data-handling capabilities of these agencies, if the agencies will make frequent use of those data bases. So that states may make optimal use of the data banks, the updating mechanism should be changed to ensure that the data are more current than data presently available from NEDS/SAROAD.

3.1.2 Intermediate Data Handling Systems

Agencies in this category are characterized as having one or more computer software systems operational for handling emissions data and/or air quality data. These states are

generally committed to the concept of total computer system development for emissions and air quality data. Generally at least one computer systems analyst is assigned full-time to the development and operation of internal systems. The mix of in-house systems and CDHS subsystems appears to be about equal. Increasing interest is being shown for EIS or AQDHS by states in which systems for emissions data or air quality data, but not both, have been developed. This category includes agencies at all stages of systems development.

NEDS and SAROAD report formats generally parallel the formats required by these agencies. Installation of in-house computer software systems precludes extensive use of NEDS or SAROAD, except where NEDS and SAROAD can provide the data bases for the projected systems. Those state agencies that have not yet made firm commitments for system development to handle either air quality data or emissions data represent potential users of NEDS or SAROAD.

3.1.3 Advanced Data Handling Systems

Agencies in this category have most of their systems for both emissions and air quality data computerized and operational. They often use extensive telemetry systems for transmitting data from ambient air sampling networks.

These states have designed and implemented their own report formats, which may or may not be compatible with those available for NEDS and SAROAD. The data bases for these systems are established. Access to NEDS and SAROAD is not an advantage to these agencies. Any access made would be to fulfill query requirements or to obtain out-of-state data. Requests of these types are infrequent.

3.2 CANDIDATE USERS FOR ALTERNATIVE SYSTEMS

The major candidates for alternative systems for accessing NEDS and SAROAD data are projected to be the state agencies classified as basic or intermediate. States classified as advanced in systems development may access NEDS and SAROAD data, but their requirements would be marginal, since most use only

data generated within their own state boundaries. States with computerized systems for handling air quality data but not for emissions data might be candidates for a NEDS access system but not for a SAROAD access system, and vice versa. Consequently the candidate users of NEDS data may not number the same as the candidate users of SAROAD data.

Table 4 shows current estimates of the status of manual and computerized systems for handling emissions and air quality data in 55 state or equivalent agencies. The figures represent the results from the EPA "State Air Data Information Survey." The CDHS figures represent states that have made direct requests to EPA for software to implement either the Emission Inventory Subsystem (EIS) or the Air Quality Data Handling Subsystem (AQDHS) of CDHS.

Table 4. ESTIMATES OF COMPUTER STATUS FOR STATE E.I. (EMISSIONS INVENTORY) AND A.Q. (AIR QUALITY) DATA SYSTEMS

	Emissions data		Air quality data	
	No.	% of total	No.	% of total
Computer	20	36	31	56
CDHS	6	11	6	11
Manual	29	53	18	33
Total	55		55	

The states in the manual category of Table 4 are the potential users of alternative systems for accessing NEDS and SAROAD data. Projections from the "State Air Data Information Survey" indicate that at least 7 states will not computerize emissions data and 3 states will not computerize air quality data. The potential users of NEDS data would be 7 to 29 agencies, and the potential users of SAROAD data would be 3 to 18 agencies.

3.3 PROJECTED ACTIVITY FOR ALTERNATIVE SYSTEMS

The minimum frequencies with which state agencies summarize emissions data and air quality data are shown in Table 5.

Table 5. MINIMUM SUMMARY FREQUENCIES FOR E.I. AND A.Q. DATA IN STATE AGENCIES

Frequency	E.I.	A.Q.
Weekly		x
Monthly	x	x
Quarterly		x
Semi-annually	x	
Annually	x	x

Correlating the frequencies in Table 5 with the numbers of reports shown in Tables 1 and 2 allows us to project ranges of annual requests for output among anticipated users of NEDS and/or SAROAD data. Table 6 shows the range of inquiries that might be made by all the candidate users of a NEDS/SAROAD access system.

Table 6. ANNUAL RANGE OF INQUIRIES FOR E.I. AND A.Q. DATA REPORTS

Frequency	E.I. (7-29 states)	A.Q. (3-18 states)
Weekly		156-936
Monthly	84-348	36-216
Quarterly		24-144
Semi-annually	28-116	
Annually	21-87	6-36
Total per year	133-551	222-1332

The projected volumes of output for each NEDS report needed to meet minimum state reporting requirements are shown in Table 7. The assumptions are that point source listings are one page each, and emissions summaries and area source listings are five pages each. Values for both the emissions summary and area source listing must be multiplied by the total number of counties. The average number of counties for all states is taken to be 70.

The average number of point sources in NEDS is assumed to be 576/state; this value is based on the EPA report "Status of

Table 7. MINIMUM E.I. REPORT VOLUME FOR A RANGE OF 7-29 STATES

E.I. report	Volume, pages/year
Point source listing	8,000-33,000
Point source listing for selected sources	2,900-12,000
Emission summary (70 counties/state, 5 pg/county)	4,900-20,000
Area source listing (70 counties/state, 5 pg/county)	2,500-10,000
Total	18,000-75,000

the National Emissions (NEDS) and Air Quality (SAROAD) Data Banks as of November 1973." States with computer systems for emissions data were eliminated from consideration.

The projected volumes of SAROAD output needed to meet minimum state reporting requirements are shown in Table 8, which gives values for two ranges. The first range is based

Table 8. MINIMUM A.Q. REPORT VOLUME FOR A RANGE OF 3-18 STATES

A. Based on reporting stations (Nov. 1973), average 25/state

A.Q. Report	Volume/year, pages
Raw data listing	4,800-29,000
Quarterly reports	300-1,800
Quarterly frequency dist.	300-1,800
Yearly frequency dist.	75-450
Yearly report by quarters	75-450
Range: volume/year	5,600-34,000

B. Based on projected stations, average 40/state

A.Q. Report	Volume/year, pages
Raw data listing	7,700-46,000
Quarterly reports	480-2,900
Quarterly frequency dist.	480-2,900
Yearly frequency dist.	120-720
Yearly report by quarters	120-720
Range: volume/year	8,900-53,000

on the number of reporting stations designated in the "Status of the National Emissions and Air Quality Data Banks as of November 1973," and the second range represents the number of stations projected for each state in its State Implementation Plan. States with known computerized air quality data systems have been eliminated from consideration. The figures shown are based on the assumption that one page of output per sensor for each report is required, and for the states known to have manual systems an average number of stations was assumed. The actual range of volumes should be somewhere between the two ranges shown.

Tables 9 and 10 show the frequencies with which states with computerized emissions inventory or air quality systems update their data bases. Again, the figures are based on the "State Air Data Information Survey".

Table 9. UPDATE FREQUENCIES FOR COMPUTERIZED E.I. DATA HANDLING SYSTEMS (FROM "STATE AIR DATA INFORMATION SURVEY")

Frequency	No. of states	% of total
Weekly	8	50
Monthly	2	13
Semi-annually	3	19
Annually	3	19
Total	16	

Table 10. UPDATE FREQUENCIES FOR COMPUTERIZED A.Q. DATA HANDLING SYSTEMS (FROM "STATE AIR DATA INFORMATION SURVEY")

Frequency	No. of states	% of total
Weekly	12	41
Monthly	14	48
Quarterly	3	10
Total	29	

The percentage figures in Tables 9 and 10 can be applied to estimates for the rate of data generation by states with manual systems to derive an indication of probable data input activities. For deriving the input volumes for emissions data and distributing those volumes according to anticipated update frequency, the following formula was used:

$$\text{No. of sources updated per update interval} = \frac{N \times 576 \times 0.05 \times F}{I}$$

where:

N = number of potential users (7-29)

576 = average number of emission sources

0.05 = estimated update volume

F = percent/update frequency (from Table 8)

I = update intervals (e.g. weekly = 52/year, monthly = 12/year, etc.)

The anticipated updates for emissions data on alternative systems are shown in Table 11.

Table 11. VOLUME OF UPDATE DATA FOR E.I. SYSTEMS (7-29 STATES)

Frequency	No. of sources	Annual total
Weekly	2-8	100-417
Monthly	2-9	26-109
Semi-annually	18-79	38-158
Annually	38-158	38-158
Total updates/year		202-842

For deriving the input volumes for air quality data and distributing those volumes according to anticipated update frequency, the following formula was applied:

$$\text{A.Q. input volume} = \sum_{1}^{7} \frac{V \times S \times F}{I}$$

where:

V = values generated/sensor

S = number of sensors/pollutant averaging time

F = percent update frequency (from Table 9)

I = update interval (weekly - 52/year, monthly = 12/year etc.)

The anticipated updates for air quality data on alternative systems are shown in Tables 12 and 13.

Table 12. REPORTING STATIONS VOLUME OF UPDATE DATA FOR A.Q. SYSTEMS (18 STATES)

Frequency	No. of input values	Annual total
Weekly	4,000	208,000
Monthly	19,000	228,000
Quarterly	12,000	48,000
Total update values/year		484,000

Table 13. PROJECTED STATIONS VOLUME OF UPDATE DATA FOR A.Q. SYSTEMS (18 STATES)

Frequency	No. of input values	Annual total
Weekly	11,000	572,000
Monthly	54,000	648,000
Quarterly	34,000	136,000
Total update values/year		1,356,000

3.4 ALTERNATIVE SYSTEMS FOR NEDS/SAROAD ACCESS

Four alternative systems can be considered for allowing state agencies with manual systems for handling emissions or air quality data to access NEDS/SAROAD.

1. EPA Regional Office Requests - In this system, states without computer systems would meet their minimum data requirements by making batch requests through the EPA Regional Offices. State agencies are currently not using this system extensively,

and the projected impact on Regional Office manpower if they were to use this option may be an important cost consideration. This system would probably not be used by the states unless an update mechanism to provide more current output could be implemented.

2. Access to NEDS/SAROAD Data via Remote Terminal - With this system NEDS and SAROAD data would be available from EIS and AQDHS implemented on a central computer. States would have access to the computer through terminals installed in the state agency offices. Capability would be provided for the states to submit data directly to the system for edit and update. The EPA Regional Offices could access the systems quarterly or semi-annually to obtain for review data submitted during each interval from any of the states. Several computers could be used for implementing this system:
    - EPA computer at National Environmental Research Center, Research Triangle Park, North Carolina.
    - OSI (Optimal Systems, Inc.) Washington, D. C. This company is currently a government contractor providing computer services to EPA.
    - GSA INFONET System
    - BCS (Boeing Computer Services, Inc.)
3. Access to NEDS Data via Remote Terminal - The states with manual air quality data systems would fulfill their data handling requirements by making batch requests for SAROAD reports through EPA Regional Offices, or they would implement AQDHS, or they could design their own systems. The NEDS users would access the data bank through remote terminals to a central computer on which EIS had been implemented. The candidate computer systems would be EPA at Research Triangle Park, OSI, BCS, or GSA INFONET.

4. Access to SAROAD Data via Remote Terminal - The states with manual emissions data systems would fulfill their data handling requirements by making batch requests for NEDS reports through EPA Regional Offices, or they could design their own systems. The SAROAD users would access the data banks through remote terminals to a central computer on which AQDHS had been implemented. The candidate computer systems would be EPA at Research Triangle Park, OSI, BCS, or GSA INFONET.

TASK II. COST ANALYSIS

1.0 EXECUTIVE SUMMARY

This is a cost/benefit analysis of four possible systems to provide state agencies with computerized data processing of their air quality and emissions data.

Each of the four alternative systems is a computer network with one central computer and remote terminals at the participating state agencies. The four networks analyzed are: (1) the GSA INFONET (Computer Sciences Corp.); (2) the EPA/OSI system (Optimal Systems, Inc.); (3) a Research Triangle-based (EPA) Network; and (4) BCS (Boeing Computer Services, Inc.), an independent commercial network. Each of these networks is assumed to use the EPA-developed Comprehensive Data Handling System (CDHS), and in particular the Air Quality Data Handling Subsystem (AQDHS) and the Emissions Inventory Subsystem (EIS).

1.1 SYSTEM OVERVIEW

There are 55 states and territories in the United States which are required to collect, use, and submit air quality and emissions data to the National Air Data Branch (NADB). Two computerized data banks have been set up at NADB to store these data. The first is the Storage and Retrieval of Aerometric Data (SAROAD) and the second being the National Emissions Data System (NEDS). Many states process data at their own facilities using their own computers. These states set up and maintain their own data files, interpret and analyze their data, and possess a practical means of transmitting their data to regional and national levels of the EPA for inclusion in NEDS and SAROAD. Many states, however, still use manual data processing systems.

Any state agency currently without an automated system for handling air quality and emissions data may want such a capability and would like it to have the following characteristics. It must be under the direct control of the state agency insofar as its own data are concerned. Access to the system for both data input and retrieval should be via teletype terminal requiring only minimal operator training. Terminals must have card readers attached. Terminal speeds and turnaround time on retrievals are not critical. Batch retrieval of lengthy reports mailed from the central facility with overnight turnaround on the central processor is adequate for virtually all purposes. Some interactive capability is desirable for editing data inputs and querying the central facility regarding report retrievals. A capability to store historical data and various summaries of the data from the state at the central processor is required. Data summaries should be updated at least quarterly with more frequent updates being desirable. Payment for system utilization should fall entirely on the user, although initial setup costs and periodic updating of the official data base would be the responsibility of the Environmental Protection Agency (EPA).

The evaluation of the four networks proceeded in the following manner. First, a standard set of assumptions was provided to a representative from each potential network in an informal interview. The inputs required from each candidate network were similarly defined, and a list was provided to the various representatives for their response. A fifth interview was conducted with a terminal supplier to get independent inputs on this critical aspect of the system, and Mr. L. Hedgepeth of the EPA was consulted regarding the implementation costs of CDHS.

The data resulting from these interviews were then transformed to a common base of costs and benefits to enable a direct comparison to be made. Various tradeoffs were made and system sensitivities investigated.

1.2 CONCLUSIONS

The general benefit of the proposed system is to provide state agencies with computerized access to and processing of their air quality and emissions inventory data. This facilitates meeting internal reporting requirements and those mandated by the EPA. The access and processing is provided by remote terminals tied into a shared central computer facility. While each of the four networks vary in precise operating procedures, costs, and perhaps some other factors, they are all capable of performing the required function and all are generally technically equivalent.

With the exception of the Research Triangle Computer Center (RTCC), all candidates are established nationwide networks. All of them have been in business for a considerable time, and all of them are or have recently been contractors to agencies of the U.S. Government and to the EPA itself.

The BCS cost estimate is the only one that includes the capability for invoicing user states directly. The other candidate networks could invoice directly, but direct invoicing would increase the cost of the system. Current government discounts, on which the GSA (GSA discount) and OSI (EPA discount) estimates are based, would not apply if the states were invoiced directly. The alternative to direct invoicing would be the provision of user accounting data to the EPA with distribution billing of the states performed by the EPA. This mechanism is unacceptable, because it increases the cost to the EPA, and because it would neccessitate multiple handling of accounting data.

An additional consideration with the candidate systems is the contractors' current hardware capacity. Both BCS and GSA INFONET have the core capacity and the network capability required. The RTCC would need additional terminal lines to service the proposed system. OSI might be expected to reach a core capacity peak in the near future, and increased core would need to be added (this input comes from a representative of the

Management Information and Data Systems Division of the EPA). Current cost estimates do not include these possibilities.

A third problem with the analysis is the inordinately high on-line storage costs of the GSA INFONET system. These estimates are based on the total data base being on-line at all times. This would not be an acceptable way to operate the system, and the CSC representatives have indicated that with a more in-depth system definition they could anticipate revised storage estimates that would be more competitive.

The capabilities of each of these networks are summarized in Table 14.

Table 14. NETWORK BENEFIT COMPARISON

Benefit	Network			
	BCS	EPA/OSI	GSA- INFONET	RTCC
Adequate experience with networks	yes	yes	yes	yes
Experience with EPA	yes	yes	yes	yes
Established nation-wide networks	yes	yes	yes	yes
Can invoice states directly at rates quoted	yes	no	no	no
Sufficient capacity	yes	maybe	yes	maybe

The contrasting costs of the four networks are derived as follows. Each network is assumed to service 25 "typical" states distributed throughout the contiguous 48 states. The estimated costs are first-year network costs including both recurring and non-recurring costs. Table 15 summarizes the cost figures.

Any final recommendation as to which system would be most cost effective to implement should be deferred until all of the problems discussed here have been resolved. Any feasibility study should address these problems.

Table 15. NETWORK FIRST-YEAR COST COMPARISON

Cost factor	Network			
	BCS	EPA/OSI	GSA-INFONET	RTCC
Recurring Costs (Per month)				
Connect time	2,600	1,500	4,700	4,700
CPU time	4,300	1,700	1,600	940
Storage time	2,400	2,480	21,000	-
Terminal rental	5,800	5,800	5,800	5,800
Central system updates	700	700	350	350
State agency operations	5,000	5,000	5,000	5,000
Subtotal (Per month)	21,000	17,000	38,000	17,000
Non-Recurring Costs				
Software and data base conversion	7,000	7,000	330	-
Initial familiarization	25,000	25,000	25,000	25,000
Subtotal	32,000	32,000	25,000	25,000
Total First-Year Cost	284,000	236,000	480,000	229,000
Total EPA Costs	21,000	20,000	10,000	9,500
Total State Costs	263,000	216,000	470,000	220,000
Typical State Costs	9,400	10,000	37,000	8,700

2.0 TECHNICAL REPORT

This is a cost/benefit analysis of four possible systems to provide state agencies with computerized data processing of their air quality and emissions data.

Each of the four alternative systems is a computer network with one central computer and remote terminals at the participating state agencies. The four networks analyzed are: (1) the GSA INFONET (2) the EPA/OSI System (Optimal Systems, Inc.); (3) a Research Triangle-based (EPA) network; and (4) BCS (Boeing Computer Services, Inc.) an independent commercial network. Each of these networks is assumed to use the EPA-developed Comprehensive Data Handling System (CDHS), and in particular the Air Quality Data Handling Subsystem (AQDHS) and the Emissions Inventory Subsystem (EIS).

The first subsection (2.1) gives an overview of the systems being considered, delineates the requirements of each system, and generally provides the background and framework within which the cost/benefit analysis is conducted.

Subsection 2.2 contains the cost/benefit analysis itself, including a standardized list of assumptions and the inputs required for each network considered. The results of informal interviews with representatives from each network are reported as is the result of an interview with a terminal supplier and the result of another interview regarding CDHS conversion. Using these inputs each network is evaluated for its cost and suitability. Various cost tradeoffs are made and sensitivities in the results are explored.

2.1 SYSTEM OVERVIEW

These are 55 states and territories in the United States which are required to collect, use, and submit air quality and emissions data to the National Air Data Branch. Two computerized data banks, SAROAD and NEDS have been setup at NADB to manage these data. Many states process their data at their own facilities, using their own computers. These states setup and maintain their own data files, interpret and analyze their data, and possess a practical means of transmitting their data to regional and national levels of the EPA for inclusion in NEDS and SAROAD. Many states, however, still use manual data processing systems.

Table T. 16 lists those states, by EPA region, which have manual emissions inventory systems and air quality data handling systems. Note that a total of 31 states are involved; 16 have neither data base automated, 13 have an automated air quality data handling system but lack an automated emissions inventory system, and 2 have an automated emissions inventory system but lack an automated air quality data handling system.

The statistical demands in terms of data input and output that these states are expected to place on a terminal network are given in Section 3.3 of the Task 1 report (Projected Activity for Alternative Systems) and are further summarized in Section 2.2 of this report. More generally, however, a state agency currently without an automated system for handling air quality and emissions data might want such a capability and would like it to have the following characteristics. It must be under the direct control of the state agency insofar as its own data are concerned. Access to the system for both data input and retrieval should be via a teletype terminal requiring only minimal operator training. Each terminal must be equipped with a card reader. Terminal speeds and turnaround time on retrievals are not critical. Batch retrieval of lengthy reports mailed from the central facility with overnight turnaround on the central processor is adequate for most purposes.

Table 16. STATES WITH MANUAL DATA SYSTEMS

	Emission inventory		Air Quality		
				No. of sensors	
EPA Region	State	Number of point sources	State	Currently reporting	Projected
Region I	Maine	380	Maine	16	47
	New Hampshire	290	New Hampshire	33	65
	Rhode Island	165	Rhode Island	64	83
	Vermont	145	Vermont	2	22
Region II	Puerto Rico	340	Puerto Rico	13	64
	Virgin Islands	85	Virgin Islands	6	10
Region III	Washington, D.C.	110	Washington, D.C.	13	27
	West Virginia	550	West Virginia	52	64
Region IV	Florida	2150			
	Georgia	1020			
	North Carolina	3075			
	Tennessee	2030			
Region V	Minnesota	690			
Region VI	Arkansas	695			
	Louisiana	1250	Louisiana	34	34
	Oklahoma	825			
Region VII	Kansas	345			
			Iowa	30	67
Region VIII	Colorado	445			
	Montana	340			
	North Dakota	180			
	South Dakota	115	South Dakota	4	10
	Utah	100			
	Wyoming	185	Wyoming	8	13

Table 16 (Continued). STATES WITH MANUAL DATA SYSTEMS

	Emission Inventory		Air Quality		
				No. of sensors	
EPA Region	State	Number of point sources	State	Currently reporting	Projected
Region IX	American Samoa	-	American Samoa	0	2
	Arizona	260	Arizona	55	72
	Guam	12	Guam	5	6
	Hawaii	475	Hawaii	39	34
			Nevada	48	55
Region X	Alaska	100	Alaska	19	38
	Idaho	350			
Totals		16707		441	713
Averages		576		25	40

Some interactive capability is desirable for editing data inputs and querying the central facility regarding report retrievals (job status). A capability to store historical data and various summaries of the data from the state at the central processor is required. Data summaries should be updated on at least a quarterly basis with more frequent updates being desirable. Payment for system utilization should fall entirely on the user, although initial setup costs and periodic updating of the official data base would be the responsibility of the EPA.

Two approaches are considered for data storage. The first would transfer the entire NEDS and SAROAD data banks to the central facility for storage and access by the individual states. This approach is not satisfactory, because it requires that large quantities of data be stored which are never (or rarely) utilized, such as the data from states that already have computerized systems and, therefore, would not be part of the proposed network. The other alternative is to place in storage only those data which have been accumulated for a given state as that state enters the network. Storage costs could thus be reduced and the states charged a more equitable storage fee. This latter alternative is more reasonable.

The computer software to be used has already been developed. It will handle air quality data in SAROAD format and emissions data in NEDS format. It is referred to as the Comprehensive Data Handling System(CDHS) and was originally developed for use by state and local air pollution control agencies as an air quality management and technical data handling system. It includes three subsystems, but only two are of direct interest in this study. The first, Air Quality Data Handling Subsystem (AQDHS) is software for processing air quality data in SAROAD format; the other, Emissions Inventory Subsystem (EIS) is software for processing the emissions data in NEDS format. In each of the networks considered, the central computer will function as the processor for each state agency

through remote terminal access. Direct invoicing of the states would be expected to increase all of the cost estimates except the BCS estimates.

2.2 CANDIDATE NETWORK EVALUATION

A standard set of assumptions was prepared and provided to a representative from each potential network in an informal interview. These assumptions are given in subsection 2.2.1 very much as they were given to each of the candidates. The inputs required from each candidate network were similarly defined and provided to the various representatives for their response. The general input requirements are listed in subsection 2.2.2 with the specific response from each candidate summarized in subsections 2.2.2.1 through 2.2.2.4. A fifth interview was conducted with a terminal supplier to get independent inputs on this critical aspect of the system, and Mr. L. Hedgepeth of the EPA was consulted regarding the implementation costs of CDHS. The results of these interviews are given in subsections 2.2.2.5 and 2.2.2.6.

The actual comparison of network costs and the discussion of results are in subsections 2.2.4 and 2.2.5.

2.2.1 Assumptions

There are up to 31 state agencies that might be interested in a terminal network for air quality data, emissions inventory data, or both. These 31 agencies may be subdivided into three groups. Group A (16 states) requiring both air quality and emissions inventory data. Group B (13 states) requiring only emissions inventory data, and Group C (2 states) requiring only air quality data.

These agencies, for the most part, are located in the more remote and less densely populated areas of the contiguous 48 states. Although Alaska, Hawaii, and the several island territories were used to derive average usage figures, the special problems associated with these agencies are excluded from this

analysis. These problems are basically communications network problems.

Each agency will utilize a low-speed terminal with punched card input. The agency will be able to interactively request a job but the job itself will be remotely run and its output mailed to the agency later. Turnaround time is assumed to be an overnight run plus mail transit time. Data entry will require an edit listing back on both the air quality and emissions inventory systems.

Annual report generation for a typical Group A state to meet minimum requirements are:

	No. of jobs[a]	Pages of computer printout[b]
Emissions Inventory	19	2400
Air Quality	74	1850-3000
Totals (Approximate)	100	5000

[a] The emissions inventory jobs are described in Table 2; the air quality jobs are described in Table 3.

[b] Pages of computer printout are derived from Tables 19 and 20 by transforming the figures given there to a per state basis.

Table 17. ANNUAL DATA VOLUMES

Annual data input volumes for a typical Group A state are:

	No. of jobs[a]	Cards of input[b]
Emissions Inventory	15	175
Air Quality	68	10000-17000
Totals (Approximate)	100	14000

Data storage volumes of currently existing data bases are as follows:

	Volume[c]	Equivalent No. of data cards[c]
Emissions Inventory	88,000	500,000
Air Quality (data values)	31,000,000	11,250,000
Totals (Approximate)	32,000,000	12,000,000

Annual additions to the entire data bases are estimated by NADB to be approximately:

	Volume	Equivalent No. of data cards
Emissions Inventory (records)	10,000	60,000
Air Quality (data values)	11,000,000	4,000,000
Totals (Approximate)	11,000,000	4,000,000

---

[a] For emissions inventory 12 monthly updates plus 2 semi-annual inputs plus one annual update give 15 jobs; air quality assumes weekly, monthly and quarterly updates.

[b] Cards of input for emissions inventory come from Table 10 of the Technical Report on Task 1 by determining the average updates per year per state and multiplying by 6 cards per update. Air quality input volume refers to Tables 11 and 12, and uses an average to get annual volumes per state.

[c] The number of emissions inventory records currently in NEDS and the number of air quality data values in SAROAD were provided by NADB.

The current storage requirements for a "typical" state are approximately:

	Volume	Equivalent No. of data cards
Emissions Inventory (records)	1,000	7,000
Air Quality (data values)	560,000	200,000
Totals (Approximate)	560,000	210,000

In addition to the above basic data storage requirements, nine 2314 disc files of data summaries would have to be moved initially. The entire data base will be updated at least quarterly and, perhaps, monthly. It is assumed that with minimal programming only that portion of the overall data base and associated summaries which apply to states utilizing the data network need be transferred to a central computer location and stored. All necessary programming would be in COBOL.

Preliminary estimates indicate that about 10 hours of computer connect time per agency per month will be required. Typical report processing times, on an IBM 360/50 CPU, are given in Table 17 as provided by NADB.

2.2.2 Network Evaluation Input

Inputs required for each network evaluation are basically those listed in Table 19. During an informal interview, a representative of each network was asked to complete this table and/or to provide the inputs from which the values could be estimated. Any other cost factors or benefits that the network representatives felt should be included in the analysis were invited.

It was specifically requested that the following factors be considered:

1. Is the assumption of overnight turnaround reasonable? Is there an extra cost for assuring overnight turnaround? What is the maximum turnaround time without priorities? What are

Table 18. TYPICAL REPORT PROCESSING TIME

EMISSIONS INVENTORY	No. of jobs annually	No. of pages of computer printout per job	Minutes of 360/50 CPU time/page	Minutes per job	Total 360/50 CPU time Minutes
Point Source Listing	2	600	0.02	12	24
Selected Source Listing No. 1	12	5	0.06	0.3	3.6
Selected Source Listing No. 2	2	30	0.06	1.8	3.6
Emissions Summary	2	350	0.09	31.5	63
Area Source Listing	1	350	0.03	10.5	10.5
Totals	19				104.7

AIR QUALITY[a]		No. of stations Operating	No. of stations Projected		No. of stations Operating	No. of stations Projected	No. of stations Operating	No. of stations Projected
Raw Data Listings	64	25	40	0.05	1.25	2.0	80.0	128.0
Quarterly Summary Reports	4	25	40	0.055	1.375	2.2	5.5	8.8
Quarterly Frequency Distributions	4	25	40	0.08	2.0	3.2	8.0	12.8
Yearly Frequency Distributions	1	25	40	0.02	0.5	0.8	0.5	0.8
Yearly Summary By Quarters	1	25	40	0.07	1.75	2.8	1.75	2.8
Totals	74						95.75	153.2

a Operating refers to average number of stations currently reporting; Projected refers to average number of stations projected in the State Implementation Plans.

Table 19. TERMINAL NETWORK COST ESTIMATES

Cost Element	Expected Value
1. Computer Connect Time Entry Time Print Time	
2. Processing (CPU) Time	
3. Storage Time Total Data Base - On-line Off-line Typical State Data Base - On-line Off-line	
4. Terminal Rental (including maintenance)	
5. Maintenance of Duplicate Systems Conversion of Data Base Monthly Maintenance	
6. Impact on State Agencies Management Initial Training and Familiarization Monthly Manpower Costs	

your particular network priorities, what do they provide, and what do they cost?

2. What are the costs for quarterly (monthly) data base updates? What are the costs for program changes (up to 50 small changes per year)?

3. What are the costs for initial conversion from the RTP Univac 1100 system? This will involve at least 100 programs (mainly retrievals) to be compiled and an interactive query capability to be implemented, as well as the initial file conversion and a program to incorporate file updates.

4. Is the capability to invoice states directly for usage of the system available? Could the capability be implemented?

Responses to the interviews are summarized in the following sections.

2.2.2.1 BCS System - One way to implement a non-EPA user network is to contract with an existing commercial network. A number of possibilities exist, but Boeing Computer Services, Inc. (BCS) was selected as being representative. BCS was presented with the information outlined in the previous two subsections of this report, and they provided the following information.[1]

BCS suggests the use of their MAINSTREAM-TSO/RJE. This system offers on-line conversational editing and program development; batch processing and a choice of output media: printed copy, punched cards, or microfiche. The MAINSTREAM-TSO is the BCS adaptation of the Time Sharing Option of the IBM Operating System. It utilizes a purchased two-million-byte IBM System/370 Model 168 computer system located at the BCS data center in McLean, Virginia. It allows terminal users to enter, store, modify, and retrieve data in a fully conversational edit mode, and to submit programs (coded in any OS-supported programming language) for compilation and execution in a remote batch environment.

1) This subsection is based on information provided by Donald D. Davenport and Larry E. Parish of the Los Angeles Offices of BCS on 30 May and 16 July 1974.

Table 20. BCS TERMINAL NETWORK COST ESTIMATES

Cost element	Expected value
1. Computer Connect Time Entry Time Print Time	$100/user/month
2. Processing (CPU) Time	$167/user/month
3. Storage Time 15% of Total Data Base - On-line Total Data Base - Off-line Typical State Data Base - On-line Off-line	 $2,400/month $780/month $280/user month $14/user/month
4. Terminal Rental (including maintenance)	$225/user/month
5. Maintenance of Duplicate Systems Conversion of Data Base Monthly Maintenance	 $7,000 $350/month
6. Impact on State Agencies Management Initial Training & Familiarization Monthly Manpower Costs	 $1,000/user $200/user/month

Cost estimates are summarized in Table 20. The derivation of each estimate is explained below:

1. Computer connect time is defined as entry time and print time. Each is charged at $10 per hour, and assuming 10 hours connect time per user per month, the $100/month/user figure results. In the worst case, BCS estimates that the total entry and retrieval cost would be $150/month/user.

2. Using the equivalent 360/50 CPU time from Table 17 and given an estimated core size of 125K-bytes per program and 2000 disc accesses per emissions inventory program, 3500 accesses for air quality raw data listings and 2500 accesses for the other air quality programs, BCS estimated a monthly processing cost of $167 per user. This is based on their pricing algorithm and internal assumptions regarding relative processing times. (The pricing algorithm is shown as Equation 1 below. Many of the terms of this equation have their own formula for deriving the quantative factor to be entered.)

Equation 1

BCS Price = Service Factor x (CPU time + disc and tape I/O + core usage + tape drives + disc spindles + job schedular) + lines printed + multiple parts + cards punched + cards read + connect charge + storage.

3. Storage costs for raw data and data summaries are derived by first assuming that the total NADB data bases were on-line full time. Then twelve million cards of raw data, or 960 million characters, would require ten Model 3330 disc packs. Three disc packs of data summaries would be required. With an on-line storage cost of $1,200/month/3330 disc pack, the monthly charge for 13 packs would be $15,600.

A more reasonable assumption would be to transfer only that portion of the NADB data base to the system that pertains to the user states. Specific volume estimates are not readily available, however. A second assumption, one that can be applied here, is that only 15 percent of the files would need be on-line at any time. The more practical estimate then becomes 0.15 x 13 x $1,200, or approximately $2,400/month for on-line storage costs.

Off-line storage costs are $60/month/disc/pack, or $780/month for a data base of 13 disc packs. Both on-line and off-line costs would be expected to be somewhat lower than the figures derived here, since the entire NADB data base would not be transferred to this system. (Estimates of how many disc packs are required for the 25 projected user states are not available. Therefore, storage costs based on the total data base were used.) A typical state using this system might require only 1/6 of a disc pack for raw data and 1/15 of a disc pack for summary data. Projected costs for on-line storage would then be $280/month. Off-line costs for a typical state would be $14/month.

4. A simple 30 character per second terminal is felt by BCS to be adequate. The rental cost for such a terminal, including installation and maintenance, is estimated by BCS to be $100 per month per user. No terminal installation costs are anticipated. The required card reader would cost an extra $125 per month, for a total terminal cost of $225 per month.

5. The cost of maintaining duplicate systems consists of an initial component and a monthly component. BCS estimates that to train one of their people (including suitable backup) will cost about $5,000. This person would be the BCS contact for all activities involving the network. It is also estimated to cost from $1,000 to $5,000 to install the data base on the BCS machine; expected value is $2,000. This includes any requisite input software development. The person assigned to the proposed terminal network would be used about 1 to 3 days each month at an anticipated cost of $200 to $500.

6. BCS suggested that each participating agency send one man and an alternate to BCS for a 1-day training session. Monthly operation for each state agency is estimated to be on the order of 10 hours per month per user at an average burdened salary of $20 per hour or $200 per month. Training costs are estimated as $500 travel and per diem plus $320 for salaries plus $180 for administrative and other costs; i.e., $1,000 per user.

7. Additional considerations included the following items:

(1) Overnight turnaround is no problem. It is in fact the lowest priority and no extra cost is associated with assuring it. Network priorities range from overnight to 10 minutes turnaround, with multiplicative factors on the computer charging unit price ranging from 0.85 to 3.

(2) BCS suggests that EPA retain programming control and enter data base updates and programming changes directly, thus obviating all costs except those included in BCS personnel costs (for coordination) and those generated by EPA in simply entering the changes.

(3) Costs of initial conversion of the CDHS from the RTP Univac system are included in the initial component of the maintenance of duplicate systems given in Item 5 of Table 20.

(4) The BCS terminal network has the capability to bill the states, and any other users such as NADB, directly for their utilization of the system at the rates quoted.

(5) A change of contractor would incur at least the costs associated with starting up the BCS system, or approximately $7,000.

(6) BCS estimates that there would be no significant impact on overall costs from reducing input or output frequencies so long as the total volume of data input and retrieval remains constant. Some reduction in CPU time could be anticipated but total cost savings are felt to be essentially negligible.

2.2.2.2 EPA/OSI System[1] - Optimal Systems, Inc.(OSI) currently is an EPA contractor. EPA Regional Offices nationwide access the OSI system for various applications. Some Regional Offices, for example, edit SAROAD data on the OSI system prior to making data submittals to NADB. Regional Offices also are using the OSI facility for modeling applications and for MARK IV applications written within the Regional Offices. Under the EPA contract OSI has the capability to provide states with access to emissions inventory and air quality data. OSI uses IBM 360/65 computers connected to a national communications network for input, output, and the proposed network. It does not, however, include a mechanism for direct billing to state users at rates that reflect current EPA discounts. This mechanism would have to be developed. The subject is treated in greater detail in the conclusion of this report.

---

* This subsection is based largely on information provided by Susan Falkson of the Management Information and Data Systems Division of the EPA in an interview on 11 June 1974.

Cost estimates are summarized in Table 21. The derivation of each estimate is explained below:

1. Connect time, consisting of entry time and print time is assumed to be 10 hours per month. From a printout of EPA/OSI Contract History for Cost, Usage, and Statistics, the unit cost of current hours (including toll-free telephone service) is given as $5.74 for April 1974. Thus, 10 hours of connect time would cost $57.40.

2. Processing time charges come from the same report alluded to above and again are for April 1974. The overnight charges are given as $375 per hour. Using the annual processing time given in the assumptions, Table 17, and assuming processing rates are twice as fast as for the IBM 360/50, the monthly charge would be from $52/month to $67/month using operating and projected air quality stations, respectively. The higher figure is used in Table 21.

3. Storage time costs are estimated from the following direct inputs. On-line storage costs were given independently as 17 cents per 3330 track per month or $1,240 per month for an entire disc pack. It was previously estimated that ten such packs would be required for the total data base and three additional packs for data summaries. Again, these figures reflect the total NADB data base storage requirements. Assuming 15 percent of the files on-line at any given time would give a projected on-line storage cost of $2,480 per month as opposed to $16,120 per month if all files were on-line continuously. Off-line storage costs are $48 per month per pack; assuming 13 packs, this is $624 per month. A typical state in the network would require about 1/6 disc pack to satisfy current requirements for raw data - an on-line storage cost of $207 per month and an off-line cost of $5 per month. On-line storage of data summaries of a typical state in the network would require about 1/15 disc pack for a total of $13 per month. Off-line storage for data summaries for the same state would be $3 per month.

Table 21. EPA/OSI TERMINAL NETWORK COST ESTIMATES

	Cost element	Expected value
1.	Computer Connect Time Entry Time Print Time	$57.40/user/month
2.	Processing (CPU) Time	$67/user/month
3.	Storage Time 15% of Total Data Base - On-line Total Data Base - Off-line Typical State Data Base - On-line Off-line	 $2,480/month $624/month $220/user/month $8/user/month
4.	Terminal Rental (including maintenance)	$225/user/month
5.	Maintenance of Duplicate Systems Conversion of Data Base Monthly Maintenance	 $7,000 $350/month
6.	Impact on State Agencies Management Initial Training and Familiarization Monthly Manpower Costs	 $1,000/user $200/user/month

4. Terminal rental figures are the same as the BCS estimates since the EPA/OSI interview agreed with BCS that the states (or at least someone else) must supply the terminals, modems, maintenance, etc. The EPA/OSI interview did not provide any direct estimates, hence those provided by BCS are used here.

5. The EPA/OSI interview did cover the conversion problem but did not provide any direct estimates of conversion costs. No problem was anticipated in the conversion process and, from what was said qualitatively, one would have no reason to assume that the EPA/OSI costs would exceed the BCS estimates. The monthly maintenance figure comes unmodified from BCS.

6. The impact on the state agency's manpower requirements also comes from the BCS estimate since none was ventured in the previously referenced interview and there seems to be no reason

why the BCS and EPA/OSI systems should result in a significantly different cost for this factor.

7. Additional considerations include the following items: items:

(1) The assumption of overnight turnaround is reasonable and there is no added cost for assuring that it is received. There are 7 priority levels in this system ranging from 15 minutes to overnight turnaround. The intermediate turnarounds are 30 minutes, 1 hour, 2 hours, 3 hours, and 4 hours. Unit cost per resource hour as of April 1974 was $1,181 for 15 minute turnaround and $375 for overnight turnaround.

(2) EPA/OSI, like BCS, suggests that EPA retain programming control and enter programming changes and data base updates directly thus obviating all costs except those generated by EPA in simply entering the changes. No estimate of this cost was attempted.

(3) The costs associated with initial conversion of the system and subsequent changing of contractors are taken from the BCS interview since no quantitative estimates of these costs were made. It was the opinion of the EPA/OSI representative that this was not a major problem, however, thus lending credence to the BCS estimate.

(4) As discussed earlier, there is currently no capability to invoice the state directly at EPA discount rates. This will be a major consideration for any final decision on a preferred network. If the capability to invoice states directly were implemented, then the processing rates quoted here may no longer apply, since these rates are based on current EPA discounts.

2.2.2.3 The GSA INFONET System[1] - General Services Administration (GSA) is authorized to establish and operate data processing centers providing services to all Federal agencies.

---

[1] This subsection is based largely on information provided by Patrick Hodges of the Information Network Division of the Computer Sciences Corporation in an interview on 27 June 1974.

There are two time-sharing systems available through GSA - the conversational GE-440 Time Sharing System (RAMUS) and the conversational/remote batch CSC (Computer Science Corp.) National Teleprocessing System (INFONET). The second of these systems is more suitable for providing states the required access to emissions inventory and air quality data. The GSA INFONET system uses Univac 1108 central processors, remote communications concentrators (RCC's), and accommodates either low- or high-speed terminals at the network user agencies.

Provision of the technical capability required by the proposed network is straightforward; however, the users cannot be charged directly for their utilization of the system at rates that reflect normal GSA discounts. If the states were invoiced directly, commercial CSC INFONET rates would apply. To qualify for the discounted rates quoted here requires that EPA be invoiced for services rendered. This invoicing can be itemized by states so that EPA could pass the charges on to the states by deducting matching funds or through some other process. This alternative would, however, place a burden on the EPA.

Cost estimates are summarized in Table 22. The derivation of each estimate is explained below:

1. Connect time, consisting of entry time and print time is assumed to be 10 hours per month. The current billing rates for low speed terminals are as shown in the following table:

Characters per second	Prime Time, Dollars/hr	Non-prime time (after 7 PM) Dollars/hr
10	5.60	3.41
15	7.44	4.03
30	9.30	4.65

The differing quantities of data, operating policies, budgets, time zones, etc., for the state agencies which will form

the network could easily result in use of any of the six combinations of terminal speeds and prime or non-prime time transmission. In order to get a single figure for use in Table 21, the six figures given above are averaged, using equal weights, resulting in a figure of $5.74 per hour. This is only coincidentally precisely the same figure given for the EPA/OSI network. There was no indication of a monthly surcharge independent of use. Communications charges to the RCC are separate, however, and are invoiced at 54 cents per month per mile. If an average of 277 miles to the RCC is assumed (see Table 23) another $123 per month is required for connect time giving the figure in Table 22.

Table 22. GSA INFONET TERMINAL NETWORK COST ESTIMATES

Cost element	Expected value
1. Computer Connect Time	$180/user/month
Entry Time	
Print Time	
2. Processing (CPU) Time	$60/user/month
3. Storage Time	
15% of Total Data Base-On-Line	$21,000/month
Total Data Base-Off-Line	$200/month
Typical State Data Base-	
On-Line	$2,350/user/month
Off-line	$4/user/month
4. Terminal Rental (including maintenance)	$225/user/month
5. Maintenance of Duplicate Systems	
Conversion of Data Base	$330
Monthly Maintenance	-0-
6. Impact on State Agencies Management	
Initial Training and Familiarization	$1,000
Monthly Manpower Costs	$200

Table 23. LENGTH OF TYPICAL GSA INFONET TELEPHONE TOLL CALL

EPA Region	State	State Capital	Concentrator city	Estimated mileage
I	Maine	Augusta	Boston	158
	New Hampshire	Concord	Boston	69
	Rhode Island	Providence	Boston	40
	Vermont	Montpelier	Boston	148
III	District of Columbia	Washington, D.C.	Washington, D.C.	0
	West Virginia	Charleston	Washington, D.C.	257
IV	Florida	Tallahassee	Atlanta	237
	Georgia	Atlanta	Atlanta	0
	North Carolina	Raleigh	Washington, D.C.	237
	Tennessee	Nashville	Memphis	207
V	Minnesota	St. Paul	Chicago	346
VI	Arkansas	Little Rock	St. Louis	296
	Louisiana	Baton Rouge	Ft. Worth	375
	Oklahoma	Oklahoma City	Ft. Worth	188
VII	Kansas	Topeka	Kansas City	59
	Iowa	Des Moines	Kansas City	188
VIII	Colorado	Denver	Denver	0
	Montana	Helena	Seattle	494
	North Dakota	Bismarck	Denver	543
	South Dakota	Pierre	Denver	405
	Utah	Salt Lake City	Denver	375
	Wyoming	Cheyenne	Denver	99
IX	Arizona	Phoenix	Los Angeles	356
	Nevada	Carson City	San Francisco	178
X	Idaho	Boise	Seattle	415
Total				5670
Average				227

2. Processing time charges are given as standard rates per Standard Resource Unit (SRU), which is approximately one-half of a CPU second. Priority 1 (72 hour turnaround) is $0.0186/SRU, Priority 4 (24-hour turnaround) is $0.0279/SRU and Priority 9 (2-hour turnaround) is $0.0992/SRU. Assuming a 24-hour turnaround and assuming that 360/50 CPU time is 20 percent slower than Univac 1108 CPU time, the monthly cost per user ranges from $47 (using operating monitoring stations) to $60 (using projected monitoring stations). The latter figure is used in Table 22.

3. Storage time costs for on-line storage were estimated directly to be $163 per day for emissions inventory data and $3,600/day for air quality data. Using 30 days per month this yields a monthly charge of $115,000. The formula for estimating on-line costs was not provided. Off-line charges were estimated by taking the total size of the data base, 12 million cards, multiplying by 80 characters per card to get 960 million characters, dividing by 100 million characters per disc pack to get 10 disc packs and multiplying by a daily charge of $0.496 per disc pack to get a daily charge of $4.96. Multiplying by 30 days per month gives the annual monthly off-line storage charge of $150. Added to this is the charge for storing three disc packs of data summaries at a charge of $0.496 per disc pack/day or $45 per month. On-line costs for the data summaries are ($13.95 per disc pack per hour) 720 hours per month (3 disc packs) = $30,000. A more realistic estimate of 15 percent of the data base on-line would cost approximately $21,000/monthly. An individual state projected for the network would require about 1/6 of a 3300 type disc pack to satisfy current requirements. This would be about $1,700 per month on-line and $2.50 off-line. The data summaries which are pertinent to the states being considered should require no more than two 3330 disc packs. Prorated costs would be about $650 per month for on-line storage and about $1 per month for off-line storage. The figures for the

typical states in this network are based on any one state leaving its data on-line continually. This assumption allows a worst case estimate.

4. Terminal costs were given as $80 per month for a Teletype Model 33 terminal using GSA discount. Since GSA discount may not apply for the proposed network, the terminal cost for this network is revised to be $225 per month per user to parallel the estimates for the other networks.

5. No problems are anticipated with transferring or maintaining the data base. Under repeated questioning it was estimated that $330 would be required to convert the data base to the CSC machine, and no subsequent costs would be charged.

6. No estimate of the impact to the states' manpower requirements was attempted. The impact should be the same as for the other networks.

7. Added considerations are given in the following items.

(1) There is no problem with overnight turnaround. Priorities range from 1 (72 hours) to 9 (2 hours). The cost for overnight (24 hour) turnaround is 1.5 times greater than the cost for a 72-hour turnaround. Two-hour turnaround is 3.5 times more expensive than a 24-hour turnaround.

(2) Again, it was suggested that EPA retain programming control and enter programmatic changes and data base updates directly, thus obviating all costs except those generated by EPA in simply entering the changes. No estimate of this cost was attempted.

(3) No costs other than those given in paragraph 5 above were associated with initial system conversion.

(4) As discussed earlier, there is currently no capability to bill the states directly at GSA discount rates.

(5) The inordinate on-line storage costs (relative to BCS and OSI) were pursued with the CSC INFONET representative, but no resolvable cause was found for the significant difference.

2.2.2.4 Research Triangle Computer Center - A basic way to implement a non-EPA user network is to provide the users with direct access to the data banks stored on the Univac 1100 Computer system at the National Environmental Research Center (NERC) in Research Triangle Park.

It was, unfortunately, not possible to pursue this option with personnel of the Data Systems Division at the NERC to the extent that would have been desirable for the purposes of this analysis. The basic input from the Research Triangle Computer Center to this study is shown in Figure 1.

Cost estimates are derived using this input and making a number of assumptions. The cost estimates are summarized in Table 24; the derivation of each estimate is explained below:

1. Connect time is again assumed to be 10 hours per month per user. From Figure 1 this would generate a charge of $40 per month. To this must be added communications charges. At least two WATS lines (at $2,200 per month per line) would be required.* These charges are prorated among all 31 potential users. Total monthly connect time, then, is $182 per user.

2. From Figure 1 processing charges are given as $300 per hour. Using the annual processing time given in the assumptions Table 17, and assuming processing rates are three times as fast for the IBM 360/50, the monthly charge would be from $28 to $36 per month using operating and projected air quality stations, respectively. The higher figure is used in Table 23.

3. Since the data bases must be maintained at the RTCC in any event it could be argued that no additional storage costs would accrue to serve as the computer center for the non-user terminal network, hence no entry is made for this cost category in Table 24.

4. Terminal rental costs were not estimated, so the $225 figure is assumed here.

* To assure reasonable service 4 to 6 lines would be required.

UNITED STATES ENVIRONMENTAL PROTECTION AGENCY
National Environmental Research Center
Research Triangle Park, North Carolina 27711

SUBJECT: Response to PRC's Recent Inquiry DATE: August 28, 1974

FROM: J. Michael Steinacher
Chief, Systems Analysis and Programming Branch

TO: Gerald Nehls
National Aerometric Data Branch

Through: Harold B. Sauls
Director, Data Systems Division

We have some problems responding to all of PRC's questions. Understandably, their comparison is designed for evaluating commercial supplies of network services. We are not really able to cost various areas until we receive guidelines on any proposed charge-back system for RTP.

In the absence of a costing yardstick, we are only able to provide the following:

1. An assumption of overnight turnaround is certainly reasonable and does represent the maximum turn-around time without priorities.

2. Our facility for billing state users would at present be strictly limited to blanket billing of NADB.

3. Cost per hour of computer connect time = $4.00.

4. Cost per Univac 1110 computing hour would be approximately $300.00.

Figure 1. Research Triangle Computer Center Input.

Table 24. RTCC TERMINAL NETWORK COST ESTIMATES

Cost element	Expected value
1. Computer Connect Time Entry Time Print Time	$182/user/month
2. Processing (CPU) Time	$36/user/month
3. Storage Time 15% of Total Data Base - On-line Total Data Base - Off-line Typical State Data Base - On-line Off-line	- - - -
4. Terminal Rental (including maintenance)	$225/user/month
5. Maintenance of Duplicate Systems Conversion of Data Base Monthly Maintenance	 - -
6. Impact on State Agencies Management Initial Training and Familiarization Monthly Manpower Costs	 $1,000/user $200/user/month

5. Using the RTCC as the focal point for the non-EPA user network would require no data conversion or maintenance of duplicate systems.

6. The impact on the state agency's manpower requirements also comes from the BCS estimates since none are available from the RTCC.

7. The information provided in Figure 1 contains all other available information from this potential network.

2.2.2.5 Terminal Devices - Representatives from each candidate network were asked for estimates of monthly terminal lease costs. None, however, were prepared to discuss the subject in detail. An independent terminal supplier was consequently asked for input on terminal costs.

Terminals generally are available on a 1 to 3 year lease basis. The terminal costed here includes a built-in acoustic

coupler, an auxiliary keyboard, a form feed, and a card reader. Typical lease terms are shown in Table 25.

Table 25. TYPICAL TERMINAL LEASE CHARGES

	One year	Two years	Three years
Terminal Printer	$125.00	$110.00	$100.00
Acoustic Coupler	16.50	14.50	12.50
Auxiliary Keyboard	9.00	8.00	7.00
Form Feed	2.50	2.00	1.75
Card Reader	115.00	100.00	90.00
TOTAL (Cost per month)	$268.00	$234.50	$211.25

For the configuration discussed here, a $50 installation fee is required. An additional $30 per year maintenance fee would be assessed after the first year.

2.2.2.6 CDHS Conversion - This network would use the Comprehensive Data Handling System software already developed by the EPA. Mr. Lloyd Hedgepeth of the EPA provided input for CDHS conversion costs.

The longest conversion time required to install CDHS on a state agency computer has been 4 man-weeks. The shortest time has been 20 man-hours. An experienced programmer/analyst should be able to install CDHS within 1 man-week on any of the computers considered for this system. Typical CPU time for compilation and test runs on CDHS are shown in Table 26.

Table 26. TYPICAL CPU TIME FOR CDHS INSTALLATION ON AN IBM 360/50

Operation	CPU time, minutes
Installation	4
Copy Procedures	1
Compilations (11 programs)	16
Testing (programs plus sorts)	5
Total	66

CDHS installation costs, then, assuming 1 man-week professional effort ($1,000) and approximately 1 hour CPU time ($1,000) could be expected to be approximately $2,000.

2.2.3. Cost Review

The contrasting costs for the four networks are shown in Table 27. Each of the four networks is assumed, in turn, to service 25 "typical" states distributed throughout the remoter parts of the contiguous 48 states. This figure excludes Alaska, Hawaii, and the four island territories from the 31 jurisdictions listed in Table 16 and assumes that each of the remaining states is typical or average. The costs which are estimated are total first year network costs including both those which are recurring and non-recurring.

Table 27 summarizes the cost figures. The derivation of the individual entries are explained below. All values are rounded to three significant figures.

The connect time charges are the entries in Tables 20, 21, 22 and 24 multiplied by 26 to account for the number of assumed states plus one additional terminal at RTP connected 10 hours per month, to enable program and data base updates to be entered. The processing time charges are similarly derived.

Storage charges present a special problem. First, the size of the data base to be stored must be determined - toal or only that attributable to the 25 states - and then the proportion to be stored on-line must be decided. For this analysis only that portion of the data base and data summaries attributable to the 25 states is assumed to be stored. On-line versus off-line storage costs are shown in the following array and are derived by multiplying the appropriate values in Tables 20, 21, and 22 by 25.

Network	On-line	Off-line
BCS	7,000	375
EPA/OSI	5,500	300
GSA INFONET	58,750	100

It is obviously very desirable to keep as much of the data bank off-line as much as possible because of the high on-line storage costs. Assume that, on the average, 15 percent of the data base is on-line and the rest is stored off-line. Some programming will be required to achieve this but its cost should be relatively invariant and insignificant. Then,

Table 27. NETWORK FIRST-YEAR COST COMPARISON

Cost factor	Network			
	BCS	EPA/OSI	GSA-INFONET	RTCC
Recurring Costs (Per month)				
Connect time	2,600	1,500	4,700	4,700
CPU time	4,300	1,700	1,600	940
Storage time	2,400	2,480	21,000	-
Terminal Rental	5,800	5,800	5,800	5,800
Central system updates	700	700	350	350
State agency operations	5,000	5,000	5,000	5,000
Subtotal (Per month)	21,000	17,000	38,000	17,000
Non-Recurring Costs				
Software and data base conversion	7,000	7,000	330	-
Initial familiarization	25,000	25,000	25,000	25,000
Subtotal	32,000	32,000	25,000	25,000
Total First-Year Cost	284,000	236,000	480,000	229,000
Total EPA Costs	21,000	20,000	10,000	9,500
Total State Costs	263,000	216,000	470,000	220,000
Typical State Costs	9,400	10,000	37,000	8,700

the figures of Table 27 result. No storage costs attributable to the non-EPA user network are assumed for the RTCC.

Terminal rental costs are taken from subsection 2.2.2.5 and assume the additional of a card reader. The cost is then $225 per terminal with 26 terminals in the network.

Central system update costs are taken directly from Tables 20, 21, 22, and 24 with an additional $350 per month added to each figure to account for EPA personnel charges incurred in entering the changes. The state agency monthly operations charge is taken directly from Tables 20, 21, 22, and 24, multiplied by 25 to account for the assumed number of states in the network.

Software and data base conversion is taken from the entries in Tables 20, 21, 22, and 24 as is the initial training and familiarization cost of $1,000 per state.

The subtotals and totals are direct additions with the monthly subtotals multiplied by 12 to get total recurring first year costs. The cost of one terminal, its connect and CPU time charges and the central system update costs were attributed to EPA as were the non-recurring software and data base conversion. All other costs were attributed to the states. Table 27 shows these costs and the per state cost of the various networks.

2.2.4. Evaluation of Cost Survey Results

For all candidate networks surveyed, except GSA INFONET, the operating cost estimates are sufficiently close to assume that they represent the typical costs that would probably be incurred. The discrepency with the GSA INFONET costs has not been resolved. We can only conclude that differences in CSC pricing algorithms and those used by the other candidate networks might require that a different approach in pricing would be necessary. A second explanation might be that the system was defined as being on-line continuously. Defining the system as having portions of the data base on-line only as needed would probably result in more realistic estimates for data storage. This approach is probably the one that would be used in any system feasibility study. Since no previous explanation

of the problem was given prior to the interviews with the network representatives, and because the problem was only defined in general terms, the cost estimates given can be expected to reflect only the cost range, but not specific costs for the proposed system. To derive more specific costs would require a more refined definition combined with a feasibility study. In further pursuing the proposed system costs, the considerations discussed here will be important.

1. For defining the system, the following information would be required:

(1) a system flow diagram

(2) definition of the specific portion of the data files to be on-line at any one time.

(3) a projection, by potential number of users in a region, of the prime time and non-prime time requirements.

(4) complete file descriptions.

2. The capacity to invoice user states directly is important. Estimates in this survey, except BCS estimates, are based on current Federal government discounts. Estimated network operation costs would be expected to increase if states were billed directly for the use of the system. If a breakdown of usage by state were provided to the EPA with subsequent billing of the states by the EPA, the EPA would incur an increased administrative cost.

3. The mechanism and frequency of transferring updates from NADB to the central system should be defined.

4. The cost of compilation for each program in the system should be supplied to obtain a better estimate of the system maintenance cost. Univac to other systems conversion costs should be considered.

5. Core storage or communications network capacity is an important consideration, since some of the representatives interviewed in this survey alluded to possible capacity problems.

6. The problem of on-line and off-line storage requirements should be better defined. The results of this preliminary survey illustrate the confusion that can result from failure to specify on-line storage requirements accurately. On-line storage would be expected to be low, since priority processing and time-sharing mode processing is not a system requirement.